

Extreme Gradient Boosting pada Peramalan Pola Curah Hujan Bulanan Kabupaten Banyuwangi

Ana Fauziah^{1*}, Hermanto Hermanto¹, Mita Akbar Sukmarini¹.

¹Universitas Bakti Indonesia, Banyuwangi, Indonesia

ana@ubibanyuwangi.ac.id*

| Received: 22/05/2024 |

Revised: -/-/

| Accepted: 29/07/2024 |

Copyright©2024 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstrak

Data meteorologi jangka panjang sangat berguna untuk mengidentifikasi tanda-tanda fenomena perubahan iklim. Fenomena tersebut mengacu pada perubahan jangka panjang kondisi fisik atmosfer bumi seperti suhu dan pola cuaca. Hal tersebut berdampak sangat besar, terutama di Banyuwangi yang merupakan salah satu wilayah produksi beras terbesar di Jawa Timur. Memprediksi tren curah hujan bulanan penting untuk mengantisipasi kegagalan panen akibat cuaca ekstrem dan bencana alam seperti banjir dan tanah longsor. Penelitian ini menggunakan parameter cuaca pada skala global seperti suhu, hujan, penguapan, kelembaban permukaan dan tekanan permukaan laut, sedangkan untuk informasi skala lokal menggunakan data curah hujan bulanan di wilayah Banyuwangi pada tahun 2011 hingga 2023. Metode *Extreme Gradient Boosting* (XGBoost) akan digunakan untuk memprediksi curah hujan bulanan dalam model *ensemble learning* berbasis pendekatan *boosting*. Secara khusus, studi ini menekankan kemampuannya untuk membangun model prediktif pada data deret waktu yang terbatas dan dampak pemisahan data terhadap performa model. Hasil terbaik ditunjukkan oleh model dengan rasio pemisahan data 1:12 atau mencakup 80% data sebagai data pelatihan. Akurasi model mencapai MAE sebesar 72.579 mm pada pelatihan dan 80.777 mm pada pengujian. Selain itu RMSE sebesar 95.940 mm pada pelatihan dan 95.775 mm pada pengujian. Hasil penelitian ini diharapkan dapat menjadi acuan untuk membangun model peramalan cuaca jangka panjang yang lebih optimal.

Kata kunci: Curah Hujan, Peramalan, Ensemble Learning, XGBoost, Banyuwangi

Abstract

Long-term meteorological data is very useful for identifying signs of climate change phenomena. The phenomenon refers to long-term changes in the physical conditions of the Earth's atmosphere, such as temperature and weather patterns. This has a huge impact, especially in Banyuwangi, which is one of the largest rice production areas in East Java. Predicting monthly rainfall trends is important to anticipate crop failures due to extreme weather and natural disasters such as floods and landslide. This research uses weather parameters on a global scale, such as temperature, rain, evaporation, surface humidity, and sea level pressure,

while for local-scale information, it uses monthly rainfall data in the Banyuwangi area from 2011 to 2023. The extreme gradient boosting (XGBoost) method will be used to predict monthly rainfall in an ensemble learning model based on the boosting approach. In particular, this study emphasizes its ability to build predictive models on limited time series and the impact of data splitting on model performance. The best results were shown by the model with a data split ratio of 1:12, or covering 80% of the data as training data. The model accuracy achieved a MAE of 72.579 mm in training and 80.777 mm in testing. In addition, the RMSE was 95.940 mm in training and 95.775 mm in testing. The results of this study are expected to be a reference for building a more optimal long-term weather forecast model.

Keywords: Rainfall, Forecast, Ensemble Learning, XGBoost, Banyuwangi

1. Pendahuluan

Wilayah Indonesia memiliki letak geografis yang istimewa yakni selain terletak pada daerah ekuator, Indonesia juga diapit oleh benua Asia dan Australia serta samudra Hindia dan samudra Pasifik. Hal ini mengakibatkan beberapa pengendali iklim seperti El Nino dan La Nina terjadi pada perairan Indonesia dan sekitarnya. Pengendali iklim lain seperti Moonson (angin muson) yang seringkali bertiup ke wilayah Indonesia mengakibatkan kondisi iklim yang semakin kompleks dan dinamis (Helmi, 2022). Curah hujan merupakan salah satu parameter iklim yang sangat berdampak pada aspek kehidupan dan berkaitan erat dengan musim daerah tropis seperti Indonesia. Curah hujan yang tinggi menunjukkan periode musim hujan dan curah hujan rendah menunjukkan periode musim kemarau.

Pengamatan pada curah hujan jangka panjang seperti pola curah hujan bulanan dapat menggambarkan awal, durasi dan akhir periode musim hujan (Bagirov dkk., 2017). Informasi tersebut sangat membantu sektor pertanian di daerah lumbung padi seperti Banyuwangi dalam pengelolaan irigasi tanaman dan penentuan musim tanam sehingga dapat meminimalkan potensi gagal panen (Nurdin, 2011). Selain itu, informasi pola curah hujan bulanan juga dapat digunakan sebagai upaya mitigasi bencana hidrologi seperti banjir dan tanah longsor (He dkk., 2024). Rumitnya perubahan sifat curah hujan yang disebabkan oleh faktor alam dan antropogenik, maka faktor fisik yang mempengaruhi karakteristik dampak curah hujan diperlukan dalam pemodelan sebagai faktor prediksi curah hujan jangka panjang (Gu dkk., 2022). Pendekatan yang berbasis luaran model iklim global, Global Climate Model (GCM), dapat digunakan dalam pengamatan pola curah hujan bulanan. Pendekatan ini menggunakan data luaran atmosfer pada grid yang berskala global di periode tertentu untuk memprediksi informasi data pada skala lokal yang lebih kecil (Najafi, 2009). Pendekatan ini cukup populer karena memiliki biaya komputasi yang murah, tidak memakan banyak waktu dan dapat digunakan di berbagai wilayah penelitian. (Salathe dkk., 2007).

Luaran GCM dirancang untuk mensimulasikan iklim baik pada masa lalu, sekarang maupun masa depan sehingga mampu memprediksi pola iklim jangka panjang (Robock dkk., 1993). Desain GCM digunakan untuk memperkirakan banyak variabel, misalnya, radiasi permukaan, kelembapan, suhu, dan curah hujan pada resolusi spasial dan temporal asli. Resolusi spasial GCM cukup besar dan bervariasi dari 100 km hingga 500 km terdiri dari sejumlah besar lapisan vertikal untuk mewakili atmosfer dan juga lautan. Untuk mempelajari dampak regional,

resolusi GCM yang kasar akan ditransformasikan ke skala lokal yang lebih kecil atau dikenal sebagai teknik downscaling (Li dkk., 2024).

Berbagai macam model prediksi curah hujan telah dilakukan menggunakan teknik berbasis luaran GCM diantaranya menggunakan pendekatan *machine learning* seperti pada penelitian Pour dkk. (2018) yang memodelkan perubahan curah hujan di Bangladesh menggunakan Support Vector Machine (SVM). Adapun Sachindra dkk. (2018) menggunakan beberapa model sekaligus yakni SVM, Artificial Neural Network (ANN), Relevant Vector Machine (RVM) dan Genetik Programming (GP) untuk analisa pengembangan model curah hujan di wilayah Australia. Teknik *machine learning* dianggap memiliki kemampuan yang lebih baik dalam mengidentifikasi pola curah hujan dibandingkan dengan teknik konvensional (Pathan dkk., 2022). Hal ini terkait dengan kemampuannya dalam menangani data dalam jumlah besar sehingga memungkinkan penggabungan dari berbagai sumber data seperti citra satelit, data radar maupun data pengamatan di lapangan (Latif dkk., 2023).

Penelitian ini akan menggunakan salah satu tipe *machine learning* yakni *ensemble learning* berupa algoritma XGBoost untuk memperkirakan pola curah hujan bulanan kabupaten Banyuwangi. Secara khusus penelitian ini bertujuan untuk mengetahui sejauh mana kemampuan model XGBoost dalam kasus regresi pada data deret waktu. Pembagian data pelatihan dan pengujian merupakan masalah yang tidak dapat dihindari dalam prediksi deret waktu. Oleh karena itu, penelitian ini akan melakukan analisis untuk menyelidiki rasio ideal dalam pembagian data pelatihan (*training set*) dan pengujian (*testing set*) terhadap performa model yang dihasilkan. Principal Component Analysis (PCA) juga akan dilibatkan pada tahap *pre-processing* untuk mereduksi variabel luaran GCM yang dihasilkan dari beberapa grid spasial.

2. Metodologi Penelitian

2.1. Data Observasi

Data luaran GCM diambil dari situs KNMI *Climate Explorer* (2022) berupa 5 parameter cuaca seperti suhu (celcius), presipitasi (mm), evaporasi (mm), kelembapan permukaan (%) dan tekanan permukaan laut (Pa).



Gambar 1. Peta daerah penelitian dengan 4 ukuran grid

Data luaran tersebut diambil pada rentang wilayah yang mencakup kawasan kabupaten Banyuwangi yakni pada koordinat 112.5°-117.5° bujur timur dan 5°-10° lintang selatan. Area tersebut terbagi menjadi 4 grid berukuran 2.5° × 2.5° sehingga akan menghasilkan total 20 variabel prediktor. Untuk variabel respon menggunakan data curah hujan bulanan yang diambil dari situs badan pusat Meteorologi, Klimatologi dan Geofisika (BMKG) kabupaten Banyuwangi (2024). Semua data baik data luaran GCM maupun data curah hujan bulanan diambil dalam kurun waktu 13 tahun yakni dari tahun 2011 hingga 2023 atau sekitar 156 bulan.

2.2. Pre-processing

Pada tahap ini langkah awal adalah melakukan pembersihan data agar menghasilkan data yang berkualitas sehingga dapat meningkatkan kualitas hasil prediksi. Tahapan dalam pembersihan data dilakukan dengan menghilangkan *outlier* dan menangani data yang hilang (*missing value*) (Siregar dkk., 2020). Tahapan pembersihan pada penelitian ini dilakukan dengan mengisi beberapa data curah hujan bulanan yang tidak terukur dengan nilai nol. Selanjutnya melakukan penskalaan variabel berupa standarisasi pada variabel prediktor. Standarisasi diperlukan agar variabel dengan nilai rentang yang lebih besar tidak mendominasi variabel lain yang memiliki rentang nilai lebih kecil serta membantu mempercepat perhitungan dalam algoritma pemodelan. Standarisasi dilakukan dengan mengubah kumpulan data sedemikian rupa hingga memiliki nilai rata-rata 0 dan standar deviasi 1. Nilai standarisasi diperoleh dengan mengurangi setiap nilai dengan rata-rata nilai kemudian membaginya dengan standar deviasi (Thara dkk., 2019)

$$X_{new} = \frac{x - \mu}{\sigma}$$

Langkah selanjutnya adalah dengan melakukan reduksi variabel menggunakan PCA. Reduksi variabel dapat meningkatkan kecepatan dan efisiensi pemodelan (Raymer dkk., 2000). Berbeda dengan teknik pemilihan variabel, reduksi variabel pada PCA bekerja dengan tetap melibatkan seluruh komponen variabel untuk diekstraksi sehingga menghasilkan jumlah komponen variabel baru yang lebih kecil. Model PCA dapat dituliskan sebagai persamaan berikut:

$$\begin{aligned} y_1 &= PC_1 = e_1'X = e_{11}X_1 + e_{12}X_2 + \dots + e_{1p}X_p \\ y_2 &= PC_2 = e_2'X = e_{21}X_1 + e_{22}X_2 + \dots + e_{2p}X_p \\ &\vdots \\ y_n &= PC_n = e_n'X = e_{n1}X_1 + e_{n2}X_2 + \dots + e_{np}X_p \end{aligned}$$

Jika terdapat p variabel maka kita dapat membuat p kombinasi linear sedemikian rupa sehingga setiap variabel saling bebas linear atau tidak berkorelasi sehingga diperoleh varian (PC_i) = $e_i' \sum e_i$ dan kovarian (PC_i, PC_k) = $e_i' \sum e_k$ dengan $i, k = 1, 2, \dots, n$ (Munandar dkk., 2022). Langkah terakhir pada tahap *pre-processing* adalah dengan membagi data menjadi data *training set* dan *testing set*. *Training set* digunakan untuk melatih algoritma XGBoost dalam membangun model prediksi. Sedangkan *testing set* digunakan untuk mengukur kinerja model setelah pelatihan. Penelitian ini berfokus pada 3 rasio pembagian data.

Table 1. Rasio pembagian data

| Rasio | Training Set | | Testing Set | |
|-------|--------------|-------------|-------------|-------------|
| | Tahun | Banyak Data | Tahun | Banyak Data |
| 1:12 | 2011-2022 | 144 | 2023 | 12 |
| 2:11 | 2011-2021 | 132 | 2022-2023 | 24 |
| 3:10 | 2011-2020 | 120 | 2021-2023 | 36 |

2.3. Membangun Model

Model prediktif pada penelitian ini menggunakan algoritma XGBoost. Algoritma XGBoost adalah model yang menggunakan beberapa pohon keputusan yang dapat digunakan untuk masalah klasifikasi dan regresi. Model-model lemah yang dihasilkan dari pohon regresi digabungkan secara berurutan dan setiap model berikutnya berusaha untuk memperbaiki kesalahan prediksi yang dilakukan oleh model sebelumnya (Chen & Guestrin, 2016). Untuk kumpulan sejumlah n data dan m variabel maka persamaan model pohon *ensemble* adalah sebagai berikut:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^k f_k(x_i), f_k \in \mathcal{F} \quad (1)$$

Pada Persamaan (1), \hat{y}_i mewakili nilai prediksi, dan f_k adalah pohon regresi sejumlah k pada himpunan pohon regresi \mathcal{F} . Selanjutnya menghitung fungsi objektif .

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega f_k \quad (2)$$

l merupakan *loss function* yang menunjukkan selisih nilai prediksi dan nilai aktual. Ωf_k menunjukkan regularisasi yang menggambarkan kompleksitas model. Regularisasi dinyatakan pada persamaan (3).

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (3)$$

Pada setiap langkah proses pelatihan XGBoost fungsi objektif akan diminalkan dengan menghasilkan model pohon regresi baru berdasarkan model yang sudah ada sebelumnya. Selanjutnya mempercepat optimasi pada fungsi objektif melalui aproksimasi orde ke 2 dan menghapus suku-suku konstanta sehingga diperoleh fungsi objektif yang lebih sederhana.

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^n \left[g_i f_i(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (4)$$

g_i dan h_i adalah gradien orde pertama dan kedua pada fungsi kesalahan. Dengan mensubstitusikan persamaan (3) pada persamaan (4) maka fungsi objektif dapat ditulis kembali menjadi persamaan berikut:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma(T) \quad (5)$$

Bobot pada tiap daun dapat dihitung menggunakan persamaan (6).

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (6)$$

Dari persamaan (5) dan (6) maka diperoleh persamaan objektif yang optimal.

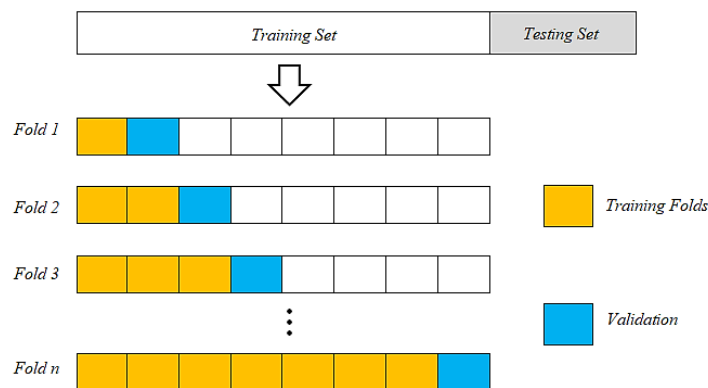
$$\tilde{\mathcal{L}}^{(t)}(q) = - \frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T$$

Fungsi objektif optimal yang menampilkan nilai prediksi pada setiap *leaf node* disajikan pada persamaan berikut:

$$\mathcal{L}_{split} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma$$

2.4. Membangun Model

Data deret waktu memiliki tantangan tersendiri dalam proses evaluasi model. Metode *cross validation* biasa kurang tepat digunakan untuk memvalidasi karakteristik data deret waktu. Hal ini dikarenakan metode tersebut tidak mempertimbangkan struktur temporal data yang bergantung pada urutan kemunculan observasi. Oleh karena itu proses validasi pada penelitian ini akan menggunakan skema *time series split*. Jumlah *split* yang akan digunakan sebanyak 11,10 dan 9 *split*. Tiap *fold* hasil *split* akan membagi *training set* menjadi data *training fold* dan data validasi dengan rentang yang berbeda. Proses *time series split* dapat dijelaskan melalui gambar berikut:



Gambar 2. Proses *time series split*

Tujuan utama dalam analisis deret waktu sering kali adalah untuk memperkirakan nilai masa depan daripada menjelaskan variasi data historis. Oleh karena itu, metrik evaluasi yang berfokus pada akurasi perkiraan dan performa model dalam struktur bergantung waktu seperti MAE dan RMSE akan digunakan dalam mengevaluasi model. Setiap *fold* akan dievaluasi menggunakan matrik evaluasi. MAE mengukur rata-rata kesalahan prediksi dengan mengambil

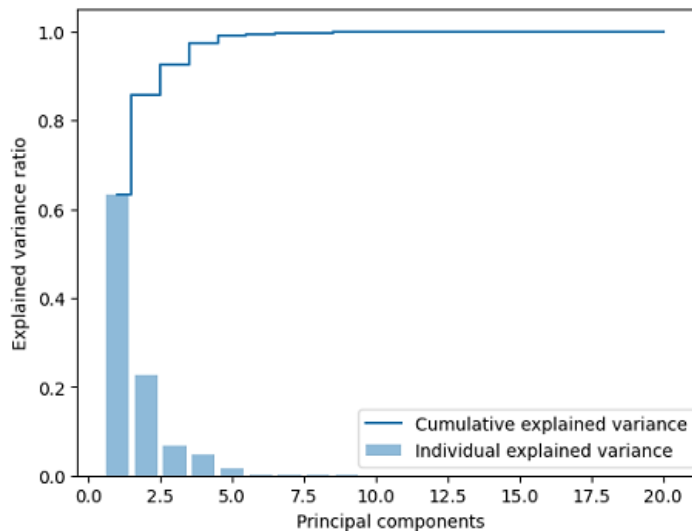
nilai absolut dari selisih antara nilai prediksi dan nilai aktual pada *training set*. Sedangkan RMSE mendefinisikan akar dari rata-rata kesalahan kuadrat. Semakin kecil nilai keduanya menunjukkan kemampuan model prediksi yang semakin baik.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \qquad RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

3. Hasil dan Pembahasan

3.1 Reduksi Dimensi

Data spasial yang digunakan pada penelitian ini menghasilkan 20 variabel prediktor yang akan direduksi menggunakan PCA. Ada beberapa cara yang digunakan untuk memilih jumlah variabel baru hasil reduksi (PC) yakni dengan memilih persentase nilai varian kumulatif lebih dari 80% atau memilih PC dengan nilai *eigen* lebih dari 1 (Vieira, 2012). Cara lain yang bisa dilakukan untuk memilih jumlah PC adalah dengan melihat bagian siku pada *scree plot* (Jhonson & Winchern, 2007).



Gambar 3. Grafik varian kumulatif

Dari Gambar 3. jumlah PC memiliki nilai *eigen* lebih dari 1 dan menunjukkan persentase varian kumulatif lebih dari 80 % sebanyak 4. Dengan pertimbangan tersebut maka variabel prediktor akan diekstraksi menjadi 4 variabel saja.

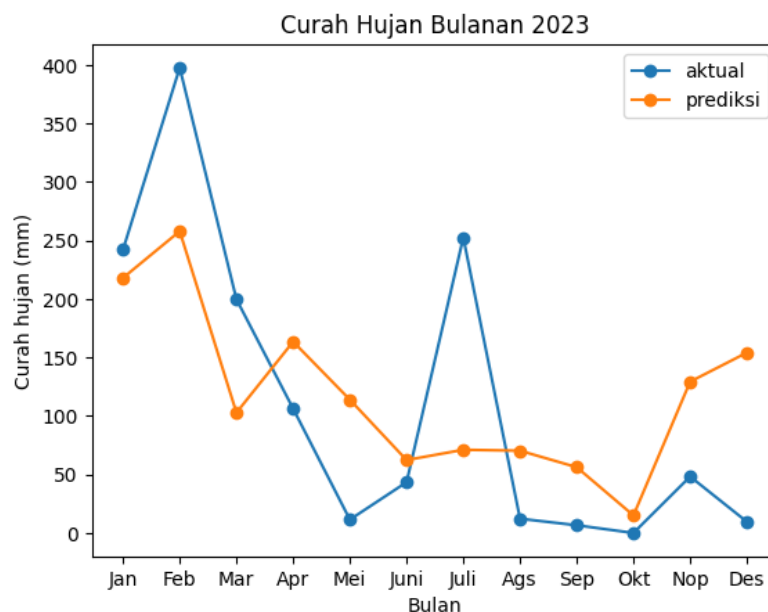
3.2 Performa Algoritma XGBoost

Algoritma XGBoost dilatih menggunakan parameter $n_estimators = 500$, $max_depth = 8$, $eta = 0.1$, $subsample = 0.8$ dan $colsample_bytree = 0.8$. Hasil evaluasi yang dilakukan pada tiap rasio pembagian data tersaji pada Table 2. Ketiga rasio pembagian data tidak menunjukkan pengaruh yang signifikan terhadap performa model yang dihasilkan. Nilai MAE dan RMSE terkecil pada *testing set* dihasilkan pada rasio 1:12. Selain itu perbedaan yang sangat kecil antara nilai evaluasi *training set* dan *testing set* pada rasio ini menunjukkan performa model yang lebih baik dibanding dengan model lainnya.

Table 2. Evaluasi Model XGBoost

| Rasio | Training Set | | Testing Set | |
|-------|--------------|----------|-------------|-----------|
| | MAE (mm) | RMSE(mm) | MAE (mm) | RMSE (mm) |
| 1:12 | 72.579 | 95.940 | 80.777 | 95.775 |
| 2:11 | 70.311 | 98.700 | 88.323 | 113.332 |
| 3:10 | 69.014 | 95.291 | 90.985 | 114.528 |

Hasil prediksi curah hujan bulanan pada *testing set* rasio 1:12 menunjukkan bahwa secara umum model mampu memprediksi pola ekstrim seperti curah hujan tinggi yang terjadi di bulan Februari dan curah hujan rendah di bulan Oktober walaupun dengan ketepatan nilai prediksi yang masih rendah.



Gambar 4. Perbandingan curah hujan bulanan aktual dan prediksi

3.3 Pembahasan

Besar kesalahan prediksi rata-rata sekitar 7.2% dari total curah hujan dalam satu tahun. Dampak signifikan dari perbedaan nilai rasio pembagian data umumnya dapat terlihat ketika bekerja pada data yang besar. Pembagian data yang tepat sangat penting untuk mencegah *overfitting* dan mendapatkan kemampuan generalisasi yang baik pada data yang belum pernah dilihat sebelumnya.

Ketidakmampuan algoritma dalam menggeneralisasi pola data juga dapat dipengaruhi oleh beberapa hal seperti ketersediaan data yang terbatas serta karakteristik dan tantangan yang unik data deret waktu. Data deret waktu biasanya menunjukkan dinamika kompleks, seperti tren, musiman dan adanya autokorelasi. Fluktuasi yang tidak tercakup secara memadai dalam kumpulan data pelatihan, menimbulkan tantangan bagi kapasitas model untuk menggeneralisasi dengan tepat.

4. Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa model XGBoost memiliki kemampuan dalam menghasilkan model prediksi data deret waktu. Model ini cukup efektif menangkap pola ketersebaran data khususnya pada kondisi ekstrem. Penggunaan 80% data sebagai data latih atau *training set* menunjukkan performa model yang lebih baik dibandingkan model dengan presentase *training set* yang lebih kecil. Penelitian selanjutnya dapat melakukan eksplorasi metode *ensemble learning* yang lain seperti teknik *stacking model*, menggunakan rekayasa fitur serta memilih parameter yang tepat melalui *hyperparameter tuning* guna meningkatkan kemampuan prediksi.

Daftar Pustaka

- Bagirov, A.M., Mahmood, A., & Barton, A. (2017). Prediction of monthly rainfall in Victoria, Australia: clusterwise linear regression approach. *Atmospheric Research*, 188, 20-29. <https://doi.org/10.1016/j.atmosres.2017.01.003>
- BMKG Banyuwangi. (2024). [Web application]. <https://www.bmkg.go.id/tag/?tag=stasiun-meteorologi-banyuwangi&lang=ID>
bmkg-sebut-fenomena-iklim-di-indonesia-kian-tak-pasti-dan-cepat-berubah-ini-sebabnya
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Gu, J., Liu, S., Zhou, Z., Chalov, S.R., & Zhuang, Q. (2022). A Stacking Ensemble Learning Model for Monthly Rainfall Prediction in the Taihu Basin, China [Special section]. *Water*, 14(3), 492.
- He, R., Zhang, L., & Chew, A.W.Z. (2024). Data-driven multi-step prediction and analysis of monthly rainfall using explainable deep learning. *Expert Systems with Applications*, 235(121160). <https://doi.org/10.1016/j.eswa.2023.121160>
- Helmi, I. (2022, Maret 23). BMKG Sebut Fenomena Iklim di Indonesia Kian Tak Pasti dan Cepat Berubah, Ini Sebabnya. *Kompas TV*. <https://www.kompas.tv/nasional/273138/>
- Jhonson, R.A. & Wichern, D.W. (2007). *Applied multivariate statistical analysis*. Pearson Prentice Hill.
- KNMI climate explorer. (2022). [Web application]. <https://climexp.knmi.nl/start.cgi?id=>
- Latif, S.D., Hazrin, N.A.B., Koo, C.H., Lin Ng, J., Chaplot, B., Huang, Y.F., El-Shafie, A., & Ahmed, A.N. (2023). Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches. *Alexandria Engineering Journal*, 82, 16-25. <https://doi.org/10.1016/j.aej.2023.09.060>
- Li, S., Xu, C., Su, M., Lu, W., Chen, Q., Huang, Q., & Teng Y. (2024). Downscaling of environmental indicators: A review. *Science of The Total Environment*, 916. <https://doi.org/10.1016/j.scitotenv.2024.170251>

- Munandar, D., Ruchjana, B.N., & Abdullah, A.S. (2022). Principal component analysis-vector autoregressive integrated (PCA-VARI) model using data mining approach to climate data in the west java region. *Jurnal Ilmu Matematika dan Terapan*, 16(1), 099-112. <https://doi.org/10.30598/barekengvol16iss1pp099-112>
- Najafi, M.R., Hamid, M., & Wherry, S.A. (2009). A procedure for statistical downscaling of precipitation with an objective method for predictor selection. *Journal of Hydrologic Engineering*. 16(8). [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000355](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000355)
- Nurdin. (2011). Antisipasi perubahan iklim untuk keberlanjutan ketahanan pangan [Srikpsi]. Universitas Negeri Gorontalo. <https://repository.ung.ac.id/karyailmiah/show/20/antisipasi-perubahan-iklim-untuk-keberlanjutan-ketahanan-pangan.html#>
- Pathan, M.S., Nag, A., & Dev, S. (2022, Juli 17-22). Efficient rainfall prediction using a dimensionality reduction method. *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing [Symposium]*. Kuala Lumpur, Malaysia. <https://doi.org/10.1109/IGARSS46834.2022.9884849>
- Pour, S.H., shahid, S., Chung, E.S., & Wang, X.J. (2018). Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh. *Atmospheric Research*, 213, 149-163. <https://doi.org/10.1016/j.atmosres.2018.06.006>
- Raymer, M.L., Punch, W.F., Goodman, E.D., Khun, L.A., & Jain, A.K. (2000). Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 4(2), 164-171. <https://corescholar.libraries.wright.edu/knoesis/9>
- Robock, A., Turco, R.P., Harwell, M.A., Ackerman, T.P., Andressen,R., Chang,H.S., & Sivakumar,M.V.K. (1993). Use of general circulation model output in the creation of climate change scenarios for impact analysis. *Climatic Change*, 23, 293-335. <https://link.springer.com/article/10.1007/BF01091621>
- Sachindra, D.A., Ahmed, K., Rashid, M.M., Shahid, S., & Perera, B.J.C (2018). Statistical downscaling of precipitation using machine learning techniques. *Atmospheric Research*, 212, 240-248. <https://doi.org/10.1016/j.atmosres.2018.05.022>
- Salathe, E. P., Jr., Mote, P. W., & Wiley, M. W. (2007). Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States pacific northwest. *Int. J. Climatol*, 27(12), 1611–1621. <https://doi.org/10.1002/joc.1540>
- Siregar, A.M., Tukino, Faisal, S., Fauzi, A., & Kadori, I. (2020). Klasifikasi untuk prediksi cuaca menggunakan esemble learning. *Jurnal Pengkajian dan Penerapan Teknik Informatika*, 13, 138-147. DOI:<https://doi.org/10.33322/petir.v13i2.998>
- Thara, D.K. Prema, P.S., & Xiong, F. (2019). Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques. *Pattern Recognition Letters*, 128, 544-550. <https://doi.org/10.1016/j.patrec.2019.10.029>
- Vieira, V.M. (2012). Permutation tests to estimate significances on Principal Components Analysis. *Computational Ecology and Software*, 2, 103-123.

[http://www.iaees.org/publications/journals/ces/articles/2012-2\(2\)/permutation-tests-to-estimate-significances.pdf](http://www.iaees.org/publications/journals/ces/articles/2012-2(2)/permutation-tests-to-estimate-significances.pdf)